

INFORMATION DISSEMINATION IMPLEMENTATION PLAN, FY 2005-2006

U.S. Government Printing Office
Office of Information Dissemination

Washington, D.C.

This document is located on *GPO Access* at
http://www.gpoaccess.gov/su_docs/fdlp/pubs/IDPlan.pdf

Comments on this document may be sent to Judith C. Russell, Managing
Director, Information Dissemination (Superintendent of Documents), at
<mailto:jrussell@gpo.gov>

INFORMATION DISSEMINATION IMPLEMENTATION PLAN, FY 2005-2006

TABLE OF CONTENTS

INTRODUCTION

EXECUTIVE SUMMARY

Chapter 1. INGEST/ACQUISITIONS

- 1.1 Acquisitions (Pending)
- 1.2 Legacy Digitization (Pending)
- 1.3 Harvesting Digital Content (Pending)

Chapter 2. ACCESS TO OFFICIAL PUBLICATIONS

- 2.1. National Bibliography of U.S. Government Publications
- 2.2 ONIX (Pending).....
- 2.3 GPO Access (Pending)
- 2.4 Persistent Identifiers (Pending)
- 2.5 Metadata Standards (Pending)

Chapter 3. VERSION CONTROL (Pending)

Chapter 4. AUTHENTICATION (Pending)

Chapter 5. PRESERVATION SERVICES

- 5.1 National Collection of U.S. Government Publications
- 5.2 FDLP Electronic Collection

Chapter 6. SERVICES TO PARTNERS AND CUSTOMERS

- 6.1 Future Directions for the FDLP (Pending)
 - 6.1.1 Electronic Depository Libraries (Pending)
- 6.2 GPO Services to Depository Libraries (Pending)
 - 6.2.1 Training and Continuing Education (Pending)
 - 6.2.2 Consultants in the Field (Pending)
 - 6.2.3 Library Collection Management Services (Pending)
- 6.3 Contact Center Services (Pending)
- 6.4 Other Publishing Standards (ISBN/ISSN) (Pending)
- 6.5 Marketing and Publicity (Pending)
- 6.6 Bibliographic Services (Pending)

Chapter 7. DISSEMINATION SERVICES

- 7.1 Purchasing Individual Copies (Pending)
 - 7.1.1 Print on Demand (pending)
 - 7.1.2 Title Management System (Pending)
- 7.2 International Exchange Program (Pending)
- 7.3 By-Law Distribution Program (Pending)
- 7.4 Storage and Distribution of Tangible Publications (Pending)

Chapter 8. GPO ACTIVITIES

8.1 Oracle Software Implementation (Pending)

8.2 Relocation of GPO (Pending)

APPENDICES

1. Glossary
2. Acronyms Used in this Paper
3. References
4. Guiding Principles

DRAFT

INTRODUCTION

[Last Update: 3/24/05]

GPO's *Strategic Vision for the 21st Century* is to deliver Federal information products and services from a flexible digital platform. This vision is described in the December 1, 2004, vision document <http://www.gpo.gov/congressional/pdfs/04strategicplan.pdf>.

The *Information Dissemination Implementation Plan, FY 2005-2006* articulates what the Information Dissemination organization (ID) will do to carry out the vision.

ID has the primary responsibility for carrying out two of the GPO's key mission elements:

- To provide, in partnership with Federal Depository libraries, for nationwide community facilities for the perpetual, free and ready public access to the printed and electronic documents, and other information products, of the Federal government.
- To distribute, on a cost recovery basis, copies of printed and electronic documents and other government information products to the general public.

Several parts of this plan have been published previously as individual drafts, and parts of it have been extensively revised based on the comments received from the Depository Library Council to the Public Printer, professional associations, the library community, and the public during the April – November 2004 comment period.

This plan is a living document and as parts are added or modified those changes will be announced and made available for review and comment.

Comments on this plan may be sent to Judith C. Russell, Managing Director, Information Dissemination (Superintendent of Documents) at jrussell@gpo.gov.

EXECUTIVE SUMMARY

[Last Update: 3/24/05]

Within GPO, it is the Information Dissemination organization (ID) that has the primary responsibility for carrying out the second and third mission elements. While these missions are expected to remain essentially the same in the future, the introduction of digital technology has changed the ways that products and services will be created and how they will look and function to meet the ever changing needs of the Federal Government and the public users of Government information.

The transition to digital publishing and dissemination expanded the universe of U.S. Government information to which current and permanent access must be provided. While the traditional role of the depository library system to ensure permanent access to tangible publications continues, a concomitant responsibility for digital information emerged for GPO. Agencies, libraries, and the public are becoming ever more reliant upon digital information. Today the FDL P is a program in which over 85% of titles are online, and that every title in FDL P is expected to be available in digital format within five years.

The Government has an obligation to provide permanent public access to its information, and GPO carries out this responsibility for all information in scope for the FDL P. GPO's plans to carry out this obligation focus on access, preservation, version control, authentication, and services to our depository library partners and the public.

This plan is a living document and will evolve over time. As a management tool, this implementation plan articulates what Information Dissemination will do to improve public access to Government publications, and to ensure FDL P publications, whether digital or tangible, remain accessible for future generations.

OVERALL ASSUMPTIONS OF THIS PLAN

[Last Update: 12/09/04]

1. No-fee access to Government information is a right of the people.
2. Users who wish to own a personal copy of a Government publication may purchase one at a reasonable cost.
3. The Government has an obligation to provide permanent public access to its information.
4. The mandates of 44 U.S.C. establish GPO's responsibility for providing permanent public access to tangible and digital U.S. Government publications.
5. The mix of institutions and users with interests in the GPO's preservation and access services includes Federal depository libraries and their users, other information

consumers, Congress, agency producers of information, information intermediaries of various kinds within and beyond GPO and the Government.

6. GPO's costs associated with developing and maintaining the National Bibliography and the National Collection are primarily borne by the Superintendent of Documents Salaries & Expenses appropriation.

Chapter 1. INGEST/ACQUISITIONS

1.1 Acquisitions (Pending)

1.2 Legacy Digitization (Pending)

1.3 Harvesting Digital Content (Pending)

Chapter 2. ACCESS TO OFFICIAL PUBLICATIONS

2.1. ACCESS: NATIONAL BIBLIOGRAPHY OF U.S. GOVERNMENT PUBLICATIONS

[Last Update: 3/18/05]

GPO intends to develop a comprehensive and authoritative NB that fully meets the statutory requirements of 44 U.S.C. 1710-1711. By law GPO is charged with preparing and publishing a “comprehensive index of public documents,” including “every document issued or published ... not confidential in character.”

Over time, the *Monthly Catalog of U.S. Government Publications* and its online counterpart, the *Catalog of U.S. Government Publications*, traditionally the principal outputs of GPO’s Cataloging and Indexing Program, became a catalog of the publications distributed in the Federal Depository Library Program (FDLP). This erosion of comprehensiveness has resulted from several factors, including fugitive documents not available to GPO, the rise in digital publishing, and ongoing constraints on program fiscal, information technology, and human resources.

This chapter examines ways to improve the comprehensiveness of GPO’s cataloging and metadata programs, to enhance the usability of GPO’s bibliographic products and services, to operate a standards-based national library quality program, and to enhance the public’s ability to identify and locate U.S. Government publications.

It is GPO’s goal to expand its cataloging initiatives to comply fully with statutory requirements, to increase the visibility and use of Government publications, and to develop a premier destination for information searchers. This goal is consistent with, and enhances the ability to deliver on, GPO’s mission: “Keeping America Informed.” It is also consistent with the more general Government goal to improve services that facilitate public interaction with the Government.

2.1.1. BACKGROUND

Typically a national bibliography covers works published in or about a specific country and generally is compiled and published by a country's national library. However, the focus of GPO's NB is entirely upon U.S. Government unclassified publications in all formats.

Many national bibliographies attempt to document a nation's published heritage, and in doing so makes that heritage known and accessible to present and future generations. This type of national bibliography lists and describes a wide variety of publications produced in a given country, or published elsewhere but of special interest or significance to that country. The country's deposit law may form the most important acquisitions channel for the national bibliography. A full national bibliography may include books, periodicals, sound recordings, microforms, music scores, pamphlets, government documents, theses, educational kits, video recordings and digital documents. It normally provides standard cataloging information for each item listed, and may include information on forthcoming publications to permit advance ordering.

Historically, in the United States, the Library of Congress (LC) was responsible for the traditional national bibliography. LC compiled cataloging records for all types of works from contributing libraries, ultimately resulting in the production of the *National Union Catalog*, which ceased publication in the 1980's. Today the nearest approximation of a national bibliography for the United States is contained in OCLC's WorldCat online catalog. WorldCat is a worldwide union catalog created and maintained collectively by more than 45,000 member libraries in 84 countries. With millions of online records built from the bibliographic and ownership information of contributing libraries, it is the largest and most comprehensive database of its kind. WorldCat contains a significant number of non-U.S. imprints, and is therefore broader in coverage than a national bibliography.

GPO has contributed in excess of 340,000 records for U.S. Federal publications to WorldCat since 1976, and other OCLC member libraries have contributed many additional thousands, resulting in a de facto national bibliography for U.S. Government publications being embedded in WorldCat. However extensive WorldCat is as a resource, there are several concerns with it. Among these are lack of visibility of Government information to users, difficulty in retrieving only on U.S. Government publications, multiple records representing the same resource, a general lack of holding or location information from depository libraries for Government publications, variations in cataloging level and application that have occurred over time, and the lack of unrestricted, no-fee public access to WorldCat.

2.1.2. WHAT IS THE NATIONAL BIBLIOGRAPHY OF U.S. GOVERNMENT PUBLICATIONS?

GPO's National Bibliography of U.S. Government Publications (NB) is a collaborative, comprehensive catalog containing descriptions and locations of U.S. Government unclassified publications in all formats. It is not a traditional national bibliography in the sense of *Canadiana*, the national bibliography of Canada.

Functionally, the NB will be a principal output of the Integrated Library System (ILS), Ex Libris' Aleph 500. GPO will develop one or more relational databases of library cataloging and other metadata records that describe and link to U.S. Government publications in a variety of formats and locations. The NB will utilize metadata resources created by GPO, by other agencies, and by partner institutions. The ILS' metasearch function will provide the capability to search across a distributed set of metadata resources from other institutions and display a combined result set. The public face of the ILS will be GPO's Online Public Access Catalog.

Initially the NB database will consist of GPO-produced bibliographic records representing titles cataloged from 1976 to the present. Once the ILS initial implementation is completed in spring 2005, GPO plans to expand the NB database by adding pre-1976 records, records created by cooperative cataloging partners and other institutions, as well as records converted from other bibliographic or metadata formats. Enhancing the NB database with records for historical publications is a project that will span multiple fiscal years, while adding records for new publications is an ongoing effort.

2.1.3. HISTORY OF THE GPO CATALOGING PROGRAM

The Printing Act of 1895 specified two products for the description and collocation of public documents. *The Catalogue of the Public Documents of the [...] Congress and of All Departments of the Government of the United States* (Document Catalogue) fulfilled the requirement for a "comprehensive index of public documents" to be prepared "at the close of each regular session of Congress." *The Monthly Catalog of United States Government Publications* met the stipulation for a "Monthly Catalogue...which shall show the publications printed during a month, where [those publications are] obtainable, and the price thereof." Thus the original relationship was of current awareness listing (*Monthly Catalog*) and permanent, authoritative, and comprehensive catalog (*Document Catalogue*).

In 1947, the cost and perennial difficulty of compiling the *Document Catalogue* led the Superintendent of Documents to recommend to the Joint Committee on Printing that the *Catalogue* be discontinued after the edition for 1940 was completed. The *Monthly Catalog* began to serve in the capacity of a single source for bibliographic information, fulfilling the requirements of the 1895 act which stand to this day in 44 USC §§1710-1711.

From 1895 until 1976, GPO arranged and described documents according to its own standards and systems. In July, 1976 along with a major revision of the appearance and functionality of the *Monthly Catalog*, GPO catalogers began producing machine-readable cataloging records (MARC 21) conforming to *Anglo-American Cataloging Rules* (AACR), *Library of Congress Subject Headings* (LCSH), the Ohio College Library Center (later Online Computer Library Center, OCLC) bibliographic standards and formats, and other applicable national level standards. Since that time, GPO has gradually expanded the emphasis of its cataloging and indexing program from the production of a single printed publication to the creation and maintenance of a database of high-quality, standards-based

bibliographic records which are used both inside GPO and out for an expanding array of bibliographic control activities.

As of 2004, GPO bibliographic data is being used to produce an abridged print edition of the *Monthly Catalog* and its online counterpart, the *Catalog of U.S. Government Publications* (CGP) on GPO Access. GPO also uses the same legacy software to produce the annual *Serials Supplement* and the biennial *U.S. Congressional Serial Set Catalog*. In addition, GPO bibliographic data is available to individual libraries directly from GPO and from a variety of commercial sources for the creation of high quality local databases and public access catalogs.

2.1.4. KEY ASSUMPTIONS FOR THIS CHAPTER

1. The Integrated Library System assists GPO in building the NB. A 21st century NB can not be operated on GPO's legacy information technology (IT) platforms.
2. GPO will continue to contribute records to OCLC.
3. Initially, GPO's core NB database format will be AACR2 cataloging records in MARC21 format. As cataloging standards and metadata schema evolve, GPO's use of metadata schema and cataloging formats will also evolve.
4. Other bibliographic formats may be ILS/NB system inputs or outputs. Possible system input or output formats include, but are not limited to:
 - a) MARC21
 - b) ONIX
 - c) GILS
 - d) Preservation metadata
 - e) Dublin Core
 - f) COSATI
5. When GPO performs original cataloging the default level of cataloging applied to U.S. Government publications, regardless of format, will be equivalent to OCLC "Full" level. Following applicable standards, GPO will apply different levels of cataloging in some cases, such as when converting records for older materials without the described publication in hand.
6. NB cataloging records will come from multiple sources, including:
 - a) GPO original cataloging
 - b) Record imports
 - c) Conversion of pre-1976 GPO bibliographic records
 - d) COSATI-MARC conversions
 - e) Harvested metadata

- f) Metadata from content originators
 - g) Records contributed by depository libraries and other partners
 - h) Preservation metadata from legacy digitization projects
- 7. The NB will provide public access and location information for all items in GPO's portion of the National Collection.
- 8. GPO cataloging policies and procedures will be documented and communicated proactively.
- 9. GPO will continue its participation in the LC Program for Cooperative Cataloging (PCC) Programs. These include:
 - a) NACO (name authorities)
 - b) SACO (subject authorities)
 - c) CONSER (serial cataloging)
- 10. GPO will continue to assign Superintendent of Documents classification numbers to publications made available to Federal depository libraries.
- 11. No fees will be charged to search the NB or to link to publicly available data described therein.
- 12. No fees will be charged to depository libraries to download bibliographic records from the NB.
- 13. Location information in the NB includes persistent links to online resources and a mechanism to identify the depository libraries that hold tangible publications.
- 14. GPO's application of the CONSER "single record option" varies with circumstances. GPO's analysis of the implications of creating a unique bibliographic record for each manifestation, or version, of a title, is that such a practice is more consistent with content management systems such as GPO's Future Digital System.

2.1.5. PLANNING ISSUES FOR INVESTIGATION

This section enumerates issues that GPO will investigate, document, and develop specific policies and procedures to cover.

1. The ILS should produce outputs programmatically, with minimal reworking or modification.
2. GPO will change its cataloging practice for archived digital publications to set the persistent link to the archived version immediately in order to reduce the necessity for

constant verification, research, and re-establishing links when the agency version is no longer available. This change means that the record in the NB definitely describes the publication retrieved by the link, and would also facilitate the use of GPO sites in metasearch applications.

3. GPO will investigate the benefits and costs of producing additional or more detailed analytics, or the use of other metadata schema or elements which could enhance information retrieval, version control, or the provenance of digital publications.
4. GPO will investigate the benefits and costs of GPO participating in standards-setting bodies such as NISO or MARBI.
5. GPO is investigating replacing its system of persistent identifiers. It is expected that Handles will replace PURLs. Some issues that will be examined include to what extent these solutions are backwardly compatible with GPO's installed base of nearly 50,000 PURLs, and whether GPO should continue to assign persistent identifiers to digital resources beyond its direct control, such as digital publications at agency sites.
6. GPO is investigating the application of industry-standard identifiers, such as ISBNs or ISSNs for all current publications with NB records.
7. The contents of the NC will be described by standard metadata schemes appropriate for various program needs, including:
 - a) Access metadata.
 - b) Preservation metadata.
 - c) Persistent links, such PURLs, Handles, or DOI (Digital Object Identifiers).
8. GPO is monitoring developments in the cataloging and metadata arenas in anticipation of potential impacts on the NB program. Some of the developments being monitored include:
 - a) FRBR.
 - b) METS.
 - c) MODS.
 - d) PREMIS.
9. GPO is seeking the most efficient and cost-effective strategy for performing the cataloging work for the NB at the necessary levels of quality and adherence to standards, including:
 - a) GPO staff.
 - b) Depository libraries and other cooperating partners.
 - c) National and agency libraries.
 - d) Publishing agencies.

e) Contractors.

10. GPO is investigating various means of providing bibliographic records to users and institutions with which GPO has long-term relationships, including:

- a) Depository libraries.
- b) Library of Congress.
- c) Bibliographic record vendors.
- d) OCLC and other bibliographic utilities.
- e) Book industry partners.

11. GPO will develop a detailed retrospective conversion plan for the pre-1976 cataloging records.

12. Determine appropriate application of other classification systems.

2.2 Standards/ONIX (Pending)

2.3 GPO Access (Pending)

2.4 Persistent Identifiers (Pending)

2.5 Metadata Standards (Pending)

Chapter 3. VERSION CONTROL (Pending)

Chapter 4. AUTHENTICATION (Pending)

Chapter 5. PRESERVATION SERVICES

5.1 PRESERVATION: NATIONAL COLLECTION OF U.S. GOVERNMENT PUBLICATIONS

[Last Update: 3/17/05]

The National Collection of U.S. Government Publications (National Collection) supports the Federal Depository Library Program (FDLP) in its mission to ensure no-fee permanent public access to the official publications of the United States Government regardless of format.

The National Collection includes preservation and access copies of digital publications and tangible publications. The National Collection will consist of geographically-dispersed collections of tangible and digital publications, located at multiple sites, and operated by

various partners within and beyond the U.S. Government. This briefing topic is focused on the National Collection dark archives for preservation of tangible and digital publications.

CONCEPTUAL OVERVIEW OF THE NATIONAL COLLECTION

| Contents | National Collection Preservation Sites | Access Collections for Public Use |
|-----------------------|--|--|
| Tangible publications | Dark archives with preservation copies | Light archives (minimal use, active preservation), or Depository library collections (normal preservation efforts) with access copies |
| Digital publications | Dark archives with digital preservation masters | Service archives at GPO or partner sites with access copies or derivatives |

For tangible publications the dark archive(s) include preservation copies which are only used when all other avenues to obtain a copy have been exhausted. Public access is provided from depository library collection access copies, or from light archives, which have minimal use and active preservation.

Preservation copies of digital publications will be maintained in dark archive(s), and used to create access copies or derivatives for delivery from GPO service archives or partner sites.

GPO will provide online public access and other publications and services derived from the digital preservation masters and other items in the National Collection. Access copies, also referred to as “derivatives,” of the stored digital publications will be available for no-fee online use by the public as well as for print-on-demand and document delivery services. Publications in GPO portions of the National Collection will be bibliographically controlled using GPO’s Integrated Library System. The cataloging records will be part of the National Bibliography data base, provided to OCLC, and made available to requesting libraries and vendors. Ultimately, preservation metadata will be included in GPO’s Future Digital System.

5.1.2. KEY ASSUMPTIONS

1. The National Collection primarily supports the FDLP goal of no-fee permanent public access, but also supports other GPO information dissemination and preservation programs, including print-on-demand for publications sales.
2. General administration of the National Collection and management of the tangible and digital publication dark archives are inherently governmental functions.

3. Publications in the National Collection will be included in the National Bibliography of U.S. Government Publications.
4. National Collection assets will be maintained in geographically-dispersed locations. A National Collection segment may be consolidated in a single location or as a distributed collection at multiple sites which together form a single collection.
5. National Collection management will be benchmarked against the criteria for assurance developed by the Center for Research Libraries or successor documentation of best practices..
6. National Collection preservation activities for digital resources will be based on the agreement between GPO and the National Archives and Records Administration (NARA) designating GPO as an archives affiliate.
7. The National Collection includes the FDLP Electronic Collection.
8. National Collection assets in the “dark archives” are held for preservation rather than public access.
9. Access copies of the National Collection assets are publicly accessible from light archives (tangible) or service archives (digital).
10. GPO will acquire tangible copies from a variety of sources, including the transfer of portions of the FDLP Collections from depository libraries to GPO.
11. The tangible publications in the National Collection will be digitized for preservation and access. After digitization the original publication, even if disbound, will be retained and preserved in case the item must be digitized again in the future.
12. Tangible copies of “born digital” publications will be produced for the dark archive as backups for the digital publications. Tangible copies in the National Collection dark archive will, to the extent practicable, be produced on archival media.

5.1.3. SCOPE

The NC will become, over time, a comprehensive set of tangible and digital titles that will back up the tangible collections in regional depository libraries or shared repositories into which depository library collections may be consolidated in the future. It is GPO’s goal, over the next three to five years, to gather and arrange a comprehensive collection of tangible documents for preservation and digitization. Most of the already existing titles for the tangible NC will be obtained through voluntary transfers from depository libraries. New titles are being acquired by GPO as they are issued.

The NC is comprehensive and includes publications of the Federal government, which are of public interest and educational value, regardless of format. Publications classified for reasons of national security and those produced solely for administrative or operational use are excluded by law from depository distribution. If however, such publications are acquired for the NC, they will be identified by metadata and included in the NB. Since the legal scope of the GPO Cataloging and Indexing Program is broader than that of the FDLP, some publications will be included in the NC solely because they are represented in the NB. The NC may also serve as the repository for publications from future GPO business initiatives.

5.1.4 FUNDING

A major part of GPO's efforts in FY 2005 is planning for the ultimate location and management of the NC and the startup of one of the NC tangible dark archives. GPO will explore the potential for establishing contractual relationships with libraries and other organizations to house the tangible NC versus maintaining and preserving the tangible and digital collections ourselves. These decisions will be made in consultation with the library community. Work performed under contract by the Center for Research Libraries, researching the characteristics of and levels of assurance for repositories and dark archives, has informed GPO's preservation planning.

Initial expenditures in FY 2005 include the costs of transporting and storing materials that are acquired for the tangible NC, purchasing storage equipment and supplies, and investing in the necessary information technology to develop and house the digital NC materials. Once the final plan is complete, we will be able to more accurately estimate the out-year funding requirements for this project, but it is anticipated that it will cost approximately \$1.5 million per year for the next five years to manage one tangible NC site. Once the tangible NC is assembled and the legacy digitization is complete, ongoing costs will cover incremental addition of new content and maintenance of the established tangible and digital NC.

5.1.5. COLLECTION OF TANGIBLE PUBLICATIONS

The NC is intended to fulfill user information needs, expand options for access, and assure that the documentary history of the United States is permanently available.

Tangible publications in the NC include:

- The format(s) in which the publication was produced, including microfiche, maps, posters, and other publications formats.
- Microfiche produced under contract for GPO, when the source document is not available.
- Tangible copies of "born digital" publications will be produced for the dark archive as backups for the digital publications in the NC. If an access or public use copy of a NC print title is required, it will generally be reproduced from a digitized version.
- Copies of every print publication selected for the FDLP and/or the NB.
- Tangible digital publications, such as CD-ROM and DVD-ROM titles, are included in the FDLP Electronic Collection segment of the NC.

5.1.6. ACQUISITIONS SOURCES FOR TANGIBLE PUBLICATIONS

GPO will use multiple sources for acquiring current and retrospective publications for the NC:

Current Acquisitions Sources:

- Riding agency print orders for two additional copies for the NC.
- Agency mailing lists.
- Acquiring fugitive documents.
- External user or publishing agency notification mechanisms.
- Depository library transfers
- Copies of publications produced by print-on-demand

Retrospective Acquisitions Sources:

- GPO records (FDLP publications in Record Group 287) at NARA.
- External user or publishing agency notification mechanisms.
- Copies offered by Federal depository libraries.
- Copies offered by other libraries, including Federal agency libraries.
- Copies or collections from libraries leaving the FDLP.
- Agency bibliographies.
- Booksellers.

5.1.7. ACCESS

The first point of access to FDLP tangible publications preserved in the NC, as shown in Table 2.1, is through the Federal depository libraries. Users requiring access to tangible publication will rely first on local depository collections, then on collections in regional depository libraries and finally on light archives or shared repositories that may be established by the depository library community in the future. A user must exhaust all opportunities for access to a tangible resource from the collections maintained in and by Federal depository libraries before seeking access to a tangible publication in the NC dark archives. The NC dark archives are not open to the public, and have no reading rooms or other public facilities. Access to publications in the NC dark archives will be provided to a digital copy or a tangible facsimile copy.

The terms and conditions for depository libraries to obtain tangible copies of publications in the NC are yet to be determined.

5.1.8. NC Maintenance

- Tangible publications in the NC may be arranged by bar code, radio-frequency identification (RFID), accession number sequence, or successive technology for robotic retrieval.
- The NC must include provisions for growth space.
- The tangible and digital dark portions of the NC will be maintained in closed, non-public locations, outside the Washington, D.C. area.
- Security will be provided to protect assets in the NC from accidental damage or loss, malicious tampering, or similar acts.
- GPO will benchmark its long-term preservation of the copies in the dark archives against current NARA guidance and preservation standards for print, microfiche and digital materials.

5.1.9. Preservation

A preservation plan that encompasses all formats and media represented in the NC will be formulated within the first six months of the existence of the NC.

Acquired retrospective materials will be evaluated upon intake and given appropriate preservation treatment.

Accepted preservation guidelines and best practices will be employed, particularly when publications are digitized.

Selection of digitization format must be consistent with long-term preservation capabilities.

5.1.10. LOCATION AND SPACE

Under contract or other binding agreement, portions of the NC may be located in other Federal agency facilities, depository libraries, or other non-Governmental organizations. Such agreements will define the roles and responsibilities of each partner institution. At least initially, the agreements will be modeled after GPO's content partnership agreements. (GPO's content partnerships may be viewed at http://www.access.gpo.gov/su_docs/fdlp/partners/index.html.)

Preservation copies of tangible items in the NC will be stored in environmentally controlled, secure facilities outside the Washington, D.C. metropolitan area. Each dark archive, using compact shelving, would entail an initial space requirement estimated at 7,500 square feet. Using a "bin" system for robotic retrieval may require less space, but higher initial infrastructure investment. GPO's objective is to establish two tangible dark archives, one managed by a library partner and one managed by NARA. One of these facilities is expected to be located in the eastern half of the National and one in the western half. Geographically separate redundant facilities for the access copies of tangible publications will be developed by GPO or its partners.

5.1.11. RELATIONSHIP WITH NARA

Like all other Federal agencies, GPO has a responsibility to transfer to the National Archives and Records Administration (NARA) those publications that are scheduled as permanent records of GPO's operation. This has historically included a record set of the tangible agency publications cataloged by GPO as well as record copies of GPO's own publications such as the *Monthly Catalog*. GPO will continue to work within applicable records schedules to ensure that its records management responsibilities are fulfilled in all media and formats.

GPO and NARA have begun a discussion concerning transforming the set of FDLP tangible publications that NARA currently holds for GPO into one of the proposed NC dark archives. That would allow NARA to move that material to storage, providing greater preservation for those materials. NARA will continue to refer users to FDLP collections for tangible documents and will use the digital copies in the EC for access. GPO is working with NARA to develop procedures for the addition of publications to the NC dark archive that were not previously distributed to depository libraries. This will allow GPO to assemble comprehensive coverage of all content that should be in the FDLP, whether it was distributed at the time of publication or not.

5.2 FDLP ELECTRONIC COLLECTION

[Last Update: 3/17/05]

In 1998 GPO established the FDLP Electronic Collection (FDLP EC) in order to provide stable, ongoing access to U.S. Government digital publications. This revised plan incorporates advances in the theory, technology, and practice of managing digital collections. Much of the revision recognizes the reality of a program in which over 85% of titles are online, and that every title in FDLP will be available in digital format within five years. The FDLP EC is part of GPO's National Collection of U.S. Government Publications (NC).

GPO provides permanent public access to Government digital publications as a continuation of its historic role in providing permanent access to tangible publications in conjunction with regional depository libraries. The dual roles of the FDLP Electronic Collection are reflected in the FDLP EC dark archive(s) for preservation (a component of the National Collection of U.S. Government Publications), and the service archives (access collections for public use).

GPO manages digital publications in the FDLP EC, links users to Federal digital publications through cataloging and persistent identifiers, ensures authenticity, provides appropriate instruction and support for Collection users, and ensures continued no-fee public access to the entire range of Government information available under the auspices of the FDLP.

The EC consists of copies of digital preservation masters in dark archives and access copies (derivatives) maintained by GPO or its partners in service archives for the convenience of reference. GPO or its partners will initiate steps, whenever feasible and cost-effective, to migrate the content or refresh the operating software as necessary to make the content readily accessible to a broad spectrum of users.

GPO's permanent public access initiatives support and complement the public information missions of the Congress, NARA, the Library of Congress and the other national libraries, and other Government agencies. Success depends on the participation and cooperation of these and other constituents at various stages of the information life cycle. GPO is leading efforts to include publications in the EC, provide metadata and locator services, as well as to facilitate partnerships between agencies and other constituents for data storage, access, and preservation.

This plan describes parameters and requirements for the FDLP EC, and will:

- Define the components of the EC.
- Outline GPO's role in providing access via cataloging and metadata.
- Describe criteria and methods for building the EC.
- Provide a functional definition of official information in the EC.
- Describe considerations for preserving the collection

5.2.1 COLLECTION OVERVIEW

FDLP collections include preservation and access copies of digital publications and tangible publications. These collection components are geographically dispersed, serve different functions, and are managed according to their specific roles in the overall program for public access to government information. As shown in Table 1 (below), the National Collection of U.S. Government Publications includes dark archives for preservation of tangible and digital publications as well as access collections for public use. For its first five years, prior to the 2003 agreement with the National Archives and Records Administration (NARA), the FDLP EC was primarily operated as an access collection.

TABLE 5.2.1. FUNCTIONAL OVERVIEW OF THE FDLP ELECTRONIC COLLECTION

| EC Contents | Preservation Copies in National Collection | Permanent Public Access to: |
|---|---|--|
| Deposited or Harvested (Born Digital) Publications Created by and acquired from US Gov't content originators or their business partners | <ul style="list-style-type: none"> • Digital preservation masters in a dark archive • CDs/DVDs produced by US Govt. agencies • Print or microfiche copies produced for dark archive backup | Official, authentic access copies (derivatives) from service archives, including: <ul style="list-style-type: none"> • <i>GPO Access</i> • Partner sites |

| | | |
|---|---|--|
| Converted Content, Digital Publications Created by GPO or under partnerships with publishing agencies | Scanning process produces digital preservation masters for the dark archive | Official, authentic copies from: <ul style="list-style-type: none"> • GPO- or agency-managed service archives, or other repositories |
| Converted Content, Digital Publications Created by depository libraries or other institutions | Scanning process produces digital preservation masters for the dark archive | Authentic copies from: <ul style="list-style-type: none"> • Service archives or other repositories • Depository or other library collections |

5.2.3. KEY ASSUMPTIONS

Within the mandates of 44 U.S.C., GPO takes responsibility for access to and preservation of digital Government publications for the FDLP. Developing the FDLP EC emphasizes building content, assuring permanent access, and capitalizing on the cooperative strengths of GPO and the FDLP to build the necessary infrastructure for preservation, authentication, identification, access, retrieval, and delivery.

This plan rests on several broad principles, core values, and subsequent assumptions about the FDLP:

1. Information included in the FDLP EC is U.S. Government information published by an official source, i.e. the publishing agency or other trusted source.
2. GPO will authenticate all EC digital content. Content that cannot be authenticated will not be included in the EC.
3. In addition to being authentic, EC content may be official or unofficial depending upon provenance and chain of custody.
4. A distributed system for preserving and providing access to digital Government information will produce the greatest benefit in return for resources invested.
5. Managing and maintaining the infrastructure, including hardware, software, and content, to provide permanent public access to FDLP digital Government publications

is an inherently governmental function and the costs will be borne by the Government and its contractual partners and not the end user.

6. A digital publication described in the NB is considered part of the EC.
7. To minimize undue complexity, maintenance, and expense, proprietary client software and other products with copyright-like barriers will be avoided, but, owing to agency decisions beyond GPO's authority, may be included where unavoidable.
8. GPO supports the use of open source standards, media and formats. When necessary GPO may standardize (take input in varied formats or manifestations and transform it for ingest so that it has common attributes) digital content to facilitate access or preservation.

5.2.4. DEFINING THE FDLP ELECTRONIC COLLECTION

The FDLP Electronic Collection is a comprehensive collection of online and tangible digital U.S. Government publications, which must meet the same basic criteria as tangible publications in the program. The EC includes digital Federal Government publications that have been created at taxpayer expense and demonstrate public interest or educational value. Publications determined by their issuing agencies to be required for strictly administrative or operational purposes or for official use only, or those classified for reasons of national security, are excluded. Information and data stored in and retrieved by means of document or content management systems, dynamic databases, or otherwise not fixed such that a consistent rendering can be returned time and again, will be considered for inclusion on a case-by-case basis, working as closely as possible with the publishing agency.

GPO does not, however, distribute, archive, assign persistent identifiers to, or otherwise make accessible information which is out of scope for the FDLP. Publications that have not been declassified or released by authorities for public access are not in scope for the FDLP. Occasionally there are situations in which the persistent identifiers in GPO cataloging records link to content at non-Governmental sites, such as educational institutions. These are cases in which GPO or the publishing agency has an official agreement with the institution that manages the site.

Every attempt will be made to make the FDLP Electronic Collection as inclusive and official as possible, within the limitations of available technology to preserve ongoing access.

Digital publications selected for the EC meet the scope criteria for the FDLP, and fall into one of the following categories:

1. Core legislative and regulatory documents on *GPO Access*, such as the *Congressional Record*, *Federal Register*, and others.
2. Digital publications published or made available by GPO, within specific agreements for services between GPO and the publishing agency

3. Digital publications published and made available by their publishing agencies, which GPO identifies, describes, and links to.
4. Tangible digital Government publications, such as CD-ROM or DVD-ROM, which GPO distributes to depository libraries.
5. Converted digital publications, typically created by scanning with or without optical character recognition, by GPO and its partners. GPO's partners may include publishing agencies and other partners such as depository libraries.

5.2.5. COLLECTION OF DIGITAL PUBLICATIONS

Digital publications may be contributed to, harvested for, or converted from a tangible original for the FDLP Electronic Collection portion of the NC. Creation includes digitization activities conducted by GPO, depository libraries, or other partners. Ingested digital publications include "born digital" files from agency publishing activities as well as publications harvested from the Web. Digital publications in the NC will initially be text with accompanying graphics, and the most prevalent file types in the near term are expected to be TIF, PDF, HTML, and ASCII. In the future the NC may include video, audio, and other non-text file types. NC policy is to preserve the digital master file used to produce print-on-demand files and other derivatives.

Every new textual publication in the current stream of processing will be digitized if a digital copy is not already available. A publication that has been digitized by GPO or its partners will be represented in the NC in multiple formats, including the original format when available, the digital preservation master and one or more access file formats.

As the legacy documents are digitized, access copies will be available for search and retrieval, dissemination, or repurposing for print-on-demand and other services. GPO will coordinate digitization efforts with the library and other interested communities to establish priorities, reduce duplication of effort and ensure the use of broadly acceptable digitization standards.

The FDLP Electronic Collection, the digital portion of the NC, will be located in multiple facilities for redundancy and security. Initially the GPO secure data storage facilities are expected to be in the Washington, D.C. area, a location outside the Washington area, and the Alternative Congressional Facility.

5.2.6. OFFICIAL STATUS OF THE EC

It is GPO's goal that any publication included in the FDLP EC is authentic and official. All FDLP EC content is authentic information, meaning that the content is verified by GPO to be complete and accurate; i.e. no data was lost or altered during the processes of acquiring it, storing it, migrating it, etc. Publications which cannot be authenticated should not be included in the EC.

However, the official status of an FDLP EC publication is dependant upon provenance, the source of the digital publication. In order to be considered official, an EC digital publication must be obtained from or its data origin verified by the publishing agency or its business partner. Examples of official digital publications include:

1. Deposited content for which GPO has been directly involved in the publication process, including digital publications created by GPO from source data files used to create print publications.
2. Harvested content obtained from trusted sources, such as digital publications harvested from publishing agency Web sites.
3. Converted content, i.e. digital publications created by GPO or by partner institutions in accordance with accepted program specifications.

The EC may also include unofficial, but still authentic, digital content, for example converted content derived from print publications distributed through the FDLP. Such content, when it represents the best available “copies,” may be accepted from unofficial sources such as institutions creating digital access copies that do not conform to the accepted specifications for preservation masters. Acceptable unofficial sources also include non-Governmental Internet archives from which GPO may obtain a digital access copy. It is GPO’s goal to replace unofficial copies thus acquired with preservation quality files when an opportunity to do so occurs. GPO may use a variety of methods to identify official digital content such as Public Key Infrastructure “keys”, secure signatures, digital watermarks or tokens, and successor technologies. The provenance of an EC digital publication will be indicated in the associated metadata.

5.2.7. DISCOVERY AND ACQUISITION

When GPO catalogs or applies other metadata services to digital publications they become part of the FDLP EC. There are numerous digital Government publications that are not included in the FDLP EC because GPO has not yet brought them under bibliographic control.

This first-level collection management activity depends upon knowledge that the publications exist. Even though GPO is engaged in information discovery and Web harvesting to acquire publications for the Collection, this activity is by itself insufficient. In order to include the broadest range of publications into the FDLP, and thereby ensure current and permanent access, GPO will employ a range of strategies. These include reliance on notification from and outreach to other agencies and notification from the depository library community. Online digital publications are identified and recommended by GPO or other program stakeholders.

After evaluation of the publication, necessary contact is made with the publishing agency, a selection decision made, bibliographic control established, and a copy of the digital

publication captured for the EC. The harvested digital publications may be stored in GPO's digital archive or at an FDLP partner site.

Sources for current acquisitions include:

1. Automated Web harvesting for individual publications.
2. Manual harvesting of individual publications from agency Web sites.
3. External user or publishing agency notification mechanisms.
4. Printing source files from agencies.
5. Official partnerships

Sources for retrospective acquisitions include:

- Deposited or harvested content obtained from an official entity or partner.
- Converted content - Digital publications created by an official FDLP partner, i.e. from legacy collection digitization projects.

5.2.8. ASSURING ONGOING ACCESS

5.2.8.1. GPO'S ROLE IN DIGITAL ARCHIVING

Archiving the EC employs a distributed architecture with storage shared among multiple partners and locations. Archival servers are operated by GPO, by GPO partners, and by third parties operating under contractual arrangements. Archived content furnished by GPO is considered GPO property and must remain in the public domain. Under certain circumstances, such as a GPO partner's inability to continue to maintain a digital archive, content archived at a partner site will return to direct GPO control.

GPO employs a multi-tiered approach to preserving and providing access to digital content:

1. Under the terms of the agreement with NARA designating GPO as an archival affiliate, *GPO Access* content is considered the preservation copy, which GPO is preserving on NARA's behalf.
2. Born digital publications on agency servers may be declared by the publishing agency to be permanent. This must be documented with a written agreement that includes failsafe measures.
3. Content managed by a GPO partner other than the publishing agency (e.g., a depository library) will be documented in written agreements which include failsafe measures.
4. Publications represented in the FDLP in digital format are being archived for permanent public access in the National Collection of U.S. Government Publications.
5. The means and circumstances for backing up digital content by creating and

preserving one or more tangible versions is under consideration.

5.2.8.2. SECURITY OF THE COLLECTION

Security of digital publications has many aspects. To gain a level of trust among users, digital files should be:

1. Secure from active, malicious alteration
2. Secure from inadvertant alteration due to error, mistake, or degradation of media
3. Produced, shared, and offered for access in an environment that balances, in its policies and practices, concern with security of systems with concern for access.

The public must be able to be assured that EC content is consistently available, official, and reliable. The FDLP EC, containing preservation and access copies of digital publications, will meet the highest practicable assurance level described in the *Draft Baseline Requirements for Digital Reformatting and Delivery of Legacy Federal Document Collections* or successor documentation.

To fully realize the potential of digital media in a networked setting, a fully redundant collection infrastructure, including mechanisms for access, organization, and preservation, must be created and maintained in more than one location.

To achieve a secure environment for access and preservation, security measures should include:

1. Fully functional data stored in a secure offline environment meeting national standards for geographic separation.
2. Fully functional redundant systems of servers and other infrastructure, as well as storage, for continuity and disaster recovery.
3. Fully documented policies and plans for addressing security concerns.

5.2.8.3. OTHER AGENCIES' ROLE IN DIGITAL ARCHIVING

Another Government agency, typically the publishing agency, may enter into a content partnership to preserve a portion of the FDLP EC. In addition, GPO may enter into a partnership to expand the content available to the Federal depository libraries. In either case, as well as in "tripartite" agreements involving GPO, one or more agencies, and a library institution, the basic parameters outlined above must be represented.

5.2.8.4. NARA'S ROLE IN DIGITAL ARCHIVING

GPO's EC complements the strategic goal of the National Archives and Records

Administration (NARA) to provide the public with access to the essential evidence of our Government. In general there are important distinctions in what is collected and maintained by NARA and GPO.

With the exception of the content on *GPO Access*, the FDLP Electronic Collection is not comprised of the record copies of digital publications. GPO's principal concern is with the information content of the publication, not with the publication's value as evidence of the activities of the Government. Inclusion of an agency digital publication in the Collection is in no way intended to be a substitute for the issuing agency's disposition of that publication to NARA in accordance with a records schedule.

Like all other Federal agencies, GPO has a responsibility to transfer to the National Archives those publications that are scheduled as permanent records of GPO's operation. NARA intends to maintain digital records in a format that is independent of specific hardware or software requirements, and requires agencies to transfer such records to NARA in accordance with regulatory specifications that support that independence. It is critical for NARA to maintain the provenance of the records and other contextual information in order to document how the records were used to carry out the functions and activities of the creating entity. This contextual information enables the records to provide evidence and accountability, and must be preserved along with the content of the Government publications that are archival records.

Chapter 6. SERVICES TO PARTNERS AND CUSTOMERS

6.1 Future Directions for the FDLP (Pending)

6.1.1 Electronic Depository Libraries (Pending)

6.2 GPO Services to Depository Libraries (Pending)

6.2.1 Training and Continuing Education (Pending)

6.2.2 Consultants in the Field (Pending)

6.2.3 Library Collection Management Services (Pending)

6.3 Contact Center Services (Pending)

6.4 Other Publishing Standards (ISBN/ISSN) (Pending)

6.5 Marketing and Publicity (Pending)

6.6 Bibliographic Services (Pending)

Chapter 7. DISSEMINATION SERVICES

7.1 Purchasing Individual Copies (Pending)

7.1.1 Print on Demand (Pending)

7.1.2 Depository Library Purchase Allowances (Pending)

7.1.3 Title Management System (Pending)

7.2 International Exchange Program (Pending)

7.3 By-Law Distribution Program (Pending)

7.4 Storage and Distribution of Tangible Publications (Pending)

Chapter 8. GPO ACTIVITIES

8.1 Oracle Software Implementation (Pending)

8.2 Relocation of GPO (Pending)

APPENDIX 1: GLOSSARY FOR INFORMATION DISSEMINATION AND GPO'S FUTURE DIGITAL SYSTEM

[Date of Last Update: 03/23/05]

Access: Tools and processes associated with finding, analyzing, ordering, and retrieving CPI or BPI.

Access aids: Tools and processes associated with finding, analyzing, retrieving, and ordering CPI or BPI.

Access Content Package (ACP): The result of ingest processing; i.e., validation, authentication, version control, transformation, verification of scope, validation or assignment persistent name, and metadata generation/capture.

Access (or service) copy: A digital publication whose characteristics (for example a screen-optimized PDF file) are designed for ease or speed of access rather than preservation.

Accessibility: Making tools and content available and usable for all users including those with disabilities; the degree to which the public is able to retrieve or obtain Government publications, either through the FDL P or directly through an digital information service established and maintained by a Government agency or its authorized agent or other delivery

channels, in a useful format or medium and in a time frame whereby the information has utility.

Application Security: The protection of application data and systems against unauthorized access to or modification of information, whether in storage, processing or transit, and against the denial of service to authorized users or the provision of service to unauthorized users, including those measures necessary to detect, document, and counter such threats at the application level. See also **Security**.

Archival information package (OAIS): Content information and its associated PDI needed to preserve the content over the long term, bound together by packaging information.

Archive: A collection with related systems and services, organized to emphasize the long-term preservation of information.

Archive management - See **Preservation**.

Authentic: Describes content that is verified by GPO to be complete and unaltered when compared to the version approved or published by the Content Originator.

Authentication: Validation of a user, a computer, or some digital object to ensure that it is what it claims to be. In the specific context of the Future Digital System, the assurance that an object is as the author or issuer intended it. See also **Certification**.

Authenticity: A digital publication's identity, source, ownership and/or other attributes are verified. Authentication also connotes that any change to the publication may be identified and tracked.

Availability - The degree to which information is obtainable through an intentional or unintentional provision of information and services.

Born digital: In the Future Digital System context, digital objects, created in a digital environment, with the intention of multiple eventual output products, potentially including hard copy, electronic presentation, and digital media. Born digital object will exist in an entirely digital lifecycle; relating to a document that was created and exists only in a digital format

Browse: To explore a body of information on the basis of the organization of the collections or by scanning lists, rather than by direct searching.

Business process information: Administrative information, non-content specific information that is used within the business process and package description (PD) to support access aids and data mining.

Cataloging and indexing: Cataloging is comprised of the processes involved in constructing a catalog: describing information or documents to identify or characterize them, providing "entry points" (terms) peculiar to the information or document, e.g., author, title,

subject, and format information, by which the information can be located and retrieved. The immediate product of cataloging is bibliographic records, which are then compiled into catalogs. Indexing is the process of compiling a set of identifiers that characterize a document or other piece of information by analyzing the content of the item and expressing it in the terms of a particular system of indexing. In GPO context, cataloging and indexing is the statutory term for the processes that produce the *Catalog of U.S. Government Publications* and its indexes. In the FDSys context, the process or results of applying bibliographic control to final published versions.

Certification: Proof of verification or authority. Process associated with ensuring that a digital object is authentically the content issued by the author or issuer. The "certificate" is a mark of veracity which is in some way joined to the object itself.

Certified: Providing proof of verification of authenticity or official status.

Collaboration: Allowing for multiple authors or content sources while maintaining digital asset and document control and provenance.

Collection of Last Resort – See **National Collection of U.S. Government Publications**

Collection plan, or Collection management plan: The policies, procedures, and systems developed to manage and ensure current and permanent public access to remotely accessible digital Government publications maintained in the National Collection.

Compose- The ability to style/format content

Composition: Creating content using FDSys applications.

Content: Information presented for human understanding.

Content Analysis- Interpretation of intended context.

Content Information (OAIS): The set of information that is the primary target for preservation, composed of the data object and its RI.

Converted content: Digital content created from a tangible publication.

Dark archive (digital): The site or electronic environment wherein a second “copy” or instance of all master and derivative digital files, data, and underlying enabling code resides and is maintained, under the control of the managing organization or its proxy. The dark archive must be inaccessible to the general public. Access to the dark repository contents and resources (“lighting” the archive) is triggered only by a specified event or condition.

Dark archive (tangible): A collection of tangible materials preserved under optimal conditions, designed to safeguard the integrity and important artifactual characteristics of the archived materials for specific potential future use or uses. Eventual use of the archived

materials (“lighting” the archives) is to be triggered by a specified event or condition. Such events might include failure or inadequacy of the “service” copy of the materials; lapse or expiration of restrictions imposed on use of the archives content; effect of the requirements of a contractual obligation regarding maintenance or use; or other events as determined under the charter of the dark archives.

Data mining: Discovery method applied to large collections of data, which proceeds by classifying and clustering data (by automated means) often from a variety of different databases, then looking for associations. Specifically applied to the analysis of use and user data for GPO systems. Tools and processes for finding, aggregating, and associating BPI to enhance internal and external business efficiencies.

Deposited content: Content received from content originators in digital form.

Derivative: A new presentation of existing content optimized for access. This does not include language translation.

Device: Content delivery mechanisms for digital media, such as data storage devices (e.g., CD, DVD, etc.), wireless handheld devices, future media, and storage at user sites.

Digital media: An intermediary mechanism consisting of data storage devices to deliver content to users’ storage or display devices.

Digital object: An item stored in a digital library or other digital collection of information, consisting of data, metadata, and an identifier.

Digital signature: A cryptographic code consisting of a hash, to indicate that data has not changed, encrypted with the public key of the creator or the signature.

Dissemination: The transfer from the stored form of a digital object in a repository to the client or user.

Dissemination information package (OAIS): An information package that contains parts or all or one or more archival information packages, to be distributed to the user or consumer as requested.

Distribution: Applying GPO processes and services to a tangible publication and sending a tangible copy to depository libraries.

Document: A digital object that is the analog of a physical document, especially in terms of logical arrangement and use.

Draft: A preliminary version of content, not yet in its finalized form.

Electronic presentation: The dynamic and temporary representation of content in digital format; strongly dependent upon file format and user’s presentation device

Emulation: Replication of a computing system to process programs and data from an earlier system that is no longer available.

Existing digital: In the current situation, publications or digital objects which are produced solely for digital dissemination, such as documents on agency web sites for which there is no printed equivalent.

Faithful digital reproduction: Digital objects that are optimally formatted and described with a view to their *quality* (functionality and use value), *persistence* (long-term access), and *interoperability* (e.g. across platforms and software environments). Faithful reproductions meet these criteria, and are intended to accurately render the underlying source document, with respect to its completeness, appearance of original pages (including tonality and color), and correct (that is, original) sequence of pages. Faithful digital reproductions will support production of legible printed facsimiles when produced in the same size as the originals (that is, 1:1).

FDLP Electronic Collection, or EC: The digital Government publications that GPO holds in storage for permanent public access through the FDLP, or are held by libraries and/or other institutions operating in partnership with the FDLP. These digital publications may be remotely accessible online publications, or tangible publications such as CD-ROMs maintained in depository library collections.

FDLP partner: A depository library or other institution that stores and maintains for permanent access segments of the Collection.

Final Published Version: Content in a specific presentation and format approved by its Content Originator for release to an audience. (See also **Government Publication; Publication**).

Fixity: the quality of being unaltered (e.g. "fixity of the text" refers to text that is demonstrably unaltered).

Format: In a general sense, the manner in which data, documents, or literature are organized, structured, named, classified, and arranged. Specifically, the organization of information for storage, printing, or display. The format of floppy disks and hard disks is the magnetic pattern laid down by the formatting utility. In a document, the format include margins, font, and alignment used for text, headers, etc. In a database, the format comprises the arrangement of data fields and field names.

Format management -See **Preservation**.

Fugitive document: A U.S. Government publication that falls within the scope of the Federal Depository Library Program (FDLP), but has not been included in the FDLP. These publications include tangible products such as ink-on-paper, microforms, CD-ROM, or

DVDs. Fugitive documents most commonly occur when Federal agencies print or procure the printing of their publications on their own, without going through GPO.

Government publication: A work of the United States Government, regardless of form or format, which is created or compiled in whole or in part at Government expense, or as required by law, except that which is required for official use only, is for strictly operational or administrative purposes having no public interest or educational value, or is classified for reasons of national security.

Granularity: The degree or level of detail available within content in the system

Granularity Policy: The system shall have the ability to certify related or continuous piece of content in context

Handle System: A comprehensive system for assigning, managing, and resolving persistent identifiers, known as "handles," for digital objects and other resources on the Internet. Handles can be used as Uniform Resource Names (URNs).

Hard copy: Tangible printed content.

Harvest: The capture and processing of Web-based documents or publications preparatory to ingest into an archive.

Harvested content: Digital content within the scope of dissemination programs that is gathered from Federal agency Web sites.

Information granularity: The degree or level of detail available in an information system. With reference to authentication, the level of detail or specificity (e.g., page, chapter, paragraph, line) to which veracity can be certified.

Ingest (OAIS): The OAIS entity that contains the services and functions that accept SIPs from Producers, prepare Archival Information packages for storage, and ensure that information packages and their supporting descriptive information packages are established within OAIS. In the FDSys, ingest processing includes validation, authentication, version control, transformation, verification of scope, validation or assignment persistent name, and metadata generation/capture.

Integrity Mark: Emblem that is used to convey authentication information to users. The mark may be visible or invisible, and all content delivery methods should have associated marks.

Interoperability: The task of building coherent services for users from components that are technically different and independently managed.

Item: A specific piece of material in a digital library or collection; a single instance, copy, or manifestation.

Light archive: A collection of tangible materials preserved under optimal conditions, designed to safeguard the integrity and important artifactual characteristics of the archived materials while supporting ongoing permitted use of those materials by the designated constituents of the archives. A light archive normally presupposes the existence of a dark archive, as a hedge against the risk of loss or damage to the light archives content through permitted uses. A light archive is also distinct from regular collections of like materials in that it systematically undertakes the active preservation of the materials as part of a cooperative or coordinated effort that may include other redundant or complementary light archives.

Localized Presentation: Temporary representation of layout or structure on a users local presentation device.

Locate (discover): The organized process of finding Web-based documents or publications that are within scope for a particular collection.

Manage: In Information Technology contexts, to add, modify, or delete content.

Manifestation: Form given to an expression of a work, e.g., by representing it in digital form.

Metadata: Data that describes the content, quality, condition, or other characteristics of other data. Metadata describes how, when, and by whom information was collected, where it resides, and how it is formatted. Metadata helps locate, interpret, or manage. In current usage several types of metadata are defined: **descriptive**, which aids in locating information; **structural/technical**, which records structures, formats, and relationships; **administrative**, which records responsibility, rights, and other information for managing the information; and **preservation**, which incorporates elements of the other types specific to preserving the information for the long term.

METS (Metadata Encoding and Transmission Standard): Essentially a standard DTD (document type definition) for interpreting XML as metadata.

Migration: Preservation of digital content where the underlying information is retained but older formats and internal structures are replaced by newer.

National Collection of U.S. Government Publications, or NC: A comprehensive collection of all in-scope publications, content that should be (or should have been) in the FDLP, regardless of form or format. The NC will consist of multiple collections of tangible and digital publications, located at multiple sites, and operated by various partners within and beyond the U.S. Government.

No-fee access: There are no charges to individual or institutional users for searching, retrieving, viewing, downloading, printing, copying, or otherwise using digital publications in scope for the FDLP.

OAIS: Open Archival Information System Reference Model (ISO 14721:2003) - A reference model for an archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a designate community. The model defines functions, activities, responsibilities, and relationships within this archive, sets forth common terms and concepts, and defined component functions which serve as the basis for planning implementation.

Official: A version that has been approved by someone with authority.

Official content: Content that falls within the scope of the FDL P EC and is approved by, contributed by, or harvested from an official source in accordance with accepted program specifications

Official source: The Federal publishing agency, its business partner, or other trusted source.

ONIX (Online Information eXchange): A standard format that publishers can use to distribute electronic information about their books to wholesale, e-tail and retail booksellers, other publishers, and anyone else involved in the sale of books.

Online: A digital publication that is published at a publicly accessible Internet site.

Online dissemination: Applying GPO processes and services to an online publication and making it available to depository libraries and the public.

Permanent Public Access, or PPA: Government publications within the scope of the FDL P remain available for continuous, no-fee public access through the program.

Persistent Name: Provides permanence of identification, resolution of location, and is expected to be globally (e.g., internationally) registered, validated, and unique

Preliminary Composition: Preparatory representation of content format or structure

Presentation Device: A device that can present content for comprehension

Preservation: The activities associated with maintaining publications for use, either in their original form or in some verifiable, usable form. Preservation may also include creation of a surrogate for the original by a conversion process, wherein the intellectual content and other essential attributes of the original are retained. For digital materials, preservation includes the management of formats of information (including possible migration to newer versions), the storage environment, and the archival arrangement of information to facilitate preservation.

Preservation description information (OAIS): Information necessary for adequate preservation of content information, including information on provenance, reference, fixity, and context.

Preservation master: A copy which maintains all of the characteristics of the original publication, from which true copies can be made.

Preservation master requirement: A set of attributes for a digital object of sufficient quality to be preserved and used as the basis for derivative products and subsequent editions, copies, or manifestations. Requirements for use, users, and state/condition/format of the source of the original object need to be noted.

Preservation processes: Activities necessary to keep content accessible and usable, including **Migration**, **Refreshment**, and **Emulation**.

Print on demand (POD): Hard copy produced in a short production cycle time and typically in small quantities.

Provenance: The chain of ownership and custody which reflects the entities that accumulated, created, used, or published information. In a traditional archival sense, provenance is an essential factor in establishing authenticity and integrity.

Publication: (N) Content approved by its Content Originator for release to an audience. See also **Government publication**.

Reference tools: Finding aids, bibliographies, and other services to assist in the locating and use of information, often less formally organized than catalogs and indexes.

Refreshment: A preservation process for data extraction, cleaning and integration, and the triggering events of these activities.

Render: To transform digital information in the form received from a repository into a display on a computer screen or other presentation to a user.

Replication: Make copies of digital material for backup, performance, reliability, or preservation.

Repository: A computer system used to store digital collections and disseminate them to users.

Requirements: In system planning, a requirement describes what users want and expect according to their various needs. Requirements draw a comprehensible picture to facilitate communications between all stakeholders in the development of a system, and outline the opportunities for development of successful products to satisfy user needs.

Rich media: An electronic presentation incorporating audio, video, text, etc.

Rider: Request by GPO, agency, or Congress that adds copies to a Request or C.O. Order placed by a publishing agency or Congress.

Search: Process or activity of locating specific information in a database or on the World Wide Web. A search involves making a statement of search terms and refining the terms until satisfactory result is returned. Searching is distinct from browsing, which facilitates locating information by presenting references to information in topical collections or other logical groupings or lists.

Secondary dark archive (digital): Multiple “copies” or instances of the dark repository, maintained as assurance against the failure or loss of the original dark repository. The secondary dark repository must provide redundancy of content to the original dark repository, and the systems and resources necessary to support access to and management of that content must be fully independent of those supporting the original dark repository content.

Secondary service repository (digital): The secondary service archive is a “mirror” of the service archive, created to provide instantaneous and continuous access to all designated constituents when the access copy or service archive is temporarily disabled.

Security: The protection of systems against unauthorized access to or modification of information, whether in storage, processing or transit, and against the denial of service to authorized users or the provision of service to unauthorized users, including those measures necessary to detect, document, and counter such threats. The measures and controls, including physical controls in conjunction with management, technical and procedural controls, that ensure the confidentiality, integrity and availability of information processed and stored by a system. See also **Application Security**.

Service archive (digital): The site or electronic environment wherein the derivative, or “use,” files and metadata created from source objects (here, tangible government documents), as well as the software, systems, and hardware necessary to transmit and make those files and metadata accessible, are maintained for public display and use. The service repository contains the current and most comprehensive electronic versions of those source materials.

Shared repository: A facility established, governed, and used by multiple institutions to provide storage space and, in some instances limited service for low-use library materials, primarily paper-based materials that do not have to be readily available for consultation in campus libraries.

Storage: The functions associated with saving digital publications on physical media, including magnetic, optical, or other alternative technologies.

Storage management - See **Preservation**.

Sub-versions of content: The state of content within the style tools and prior to ingest.

Submission information package (OAIS): The information package identified by the producer for ingest into an OAIS system.

Subscription: An agreement by which a user obtains access to requested content by payment of a periodic fee or other agreed upon terms.

System: An organized collection of components that have been optimized to work together in a functional whole.

Transformation: The process, or the results of a process, of reformatting or otherwise changing the way content is digitally encoded.

Trusted content: Official content that is provided by or certified by a trusted source.

Trusted source: The publishing agency or a GPO partner that provides or certifies official FDLP content.

Unique Identifier: A character string that uniquely identifies digital objects, content packages and jobs within the system.

User: The person who uses a program, system, or collection of information to perform tasks and produce results.

Validation: A process that ensures data entered into the system conforms to standards for format, content and metadata.

Verification: The process of determining and assuring accuracy and completeness.

Version: Unique manifestation of a publication.

Version control: Relating to a specific manifestation, revision, issuance, or edition of a previously published or issued document or publication. Changes beyond an agreed upon threshold or tolerance constitute a new version. That threshold is a version trigger, and the activity of scanning for changes and activating the trigger is "version control."

Version detection: Activity of inspecting a content package for changes and responding to version triggers. Also, activity of polling the system to identify if an identical version already exists in the system.

Version identifier: Information stored in metadata that identifies version.

Version trigger: Changes beyond an agreed upon threshold or tolerance which constitute a new version.

APPENDIX 2. ACRONYMS USED IN THIS PAPER

[Last Update: 1/8/05]

- AACR2 – Anglo-American Cataloging Rules, 2nd edition.

- BIBCO - Bibliographic Record Component Program of the Program for Corporate Cataloging.
- CDS – Cataloging Distribution Service, Library of Congress
- CONSER – Cooperative program for serial cataloging.
- COSATI – Council on Scientific and Technical Information ; Cataloging format used by scientific and technical agencies.
- DC, or Dublin Core – Metadata format developed by OCLC.
- FDLP - Federal Depository Library Program.
- FLICC – Federal Library and Information Center Committee, LC.
- FRBR - Functional Requirements for Bibliographic Records.
- GILS – Government Information Locator Systems.
- IFLA - The International Federation of Library Associations and Institutions.
- IT - Information technology.
- ILS - Integrated Library System.
- ISBNs – International Standard Book Numbers.
- LC - Library of Congress.
- MARBI - Machine-Readable Bibliographic Information Committee of ALA.
- MARC – Machine-Readable Cataloging, a data encoding structure.
- METS - Metadata Encoding and Transmission Standard.
- MODS - Metadata Object Description Schema.
- NACO – Cooperative program to establish name authorities.
- NISO - National Information Standards Organization.
- NUC - *National Union Catalog*.
- OCLC – Online Computer Library Center, Dublin, OH.
- ONIX – Online Information Exchange; book industry cataloging format.
- OPAC - Online Public Access Catalog.
- PCC - Program for Cooperative Cataloging, LC.
- PREMIS - PREservation Metadata: Implementation Strategies, an OCLC working group.
- PURL – Persistent Uniform Resource Locator.
- SACO – Cooperative program to establish subject authorities.
- SuDocs - Superintendent of Documents, GPO.
- U.S.C. – *United States Code*.

APPENDIX 3. REFERENCES

[Last Update: 3/17/05]

British National Bibliography (BNB) description.

<http://www.bl.uk/services/bibliographic/natbib.html>

Canadiana: The National Bibliography of Canada.

<http://www.collectionscanada.ca/canadiana/index-e.html>

Collection of Last Resort, Revised Draft June 18, 2004

www.gpoaccess.gov/about/reports/clr0604draft.pdf

Decision Framework for Federal Document Repositories, Discussion Draft, April 12, 2004, prepared for GPO by the Center for Research Libraries,
www.access.gpo.gov/su_docs/fdlp/pubs/decisionmatrix.pdf

Defining What Government Information is to be Categorized: Statement of Requirements. (GPO, 2004) (<http://www.gpoaccess.gov/cgiwg/pdf/cgiwgroup/revMay2004.pdf>)

Draft Baseline Requirements for Digital Reformatting and Delivery of Legacy Federal Document Collections, November 3, 2004, prepared for GPO by the Center for Research Libraries, unpublished.

FDLP Guidelines for Determining Superseded Materials
http://www.access.gpo.gov/su_docs/fdlp/coll-dev/supersede.html

Format determination in GPO Cataloging Practices, April, 2004,
http://www.access.gpo.gov/su_docs/fdlp/cip/cataloging-formats.pdf

Functional Requirements for Bibliographic Records: Final Report. (International Federation of Library Associations, Cataloging Section, FRBR Review Group, 2003)
<http://www.ifla.org/VII/s13/wgfrbr/finalreport.htm>

GPO's Future Direction & Future System, October 17, 2004,
http://www.gpo.gov/news/2004/04_10_18_dlc_presentation.pdf

GPO's Future Digital System Concept of Operations, October, 2004,
http://www.gpo.gov/news/2004/ConOps_1004.pdf

Handle System Introduction, Corporation for National Research Initiatives, 2003, <http://www.handle.net/introduction.html>

The large-scale archival storage of digital objects, Jim Linden, et al, The British Library, Digital Preservation Coalition Technology Watch Series Report 04-03, February 2005,
<http://www.dpconline.org/docs/dpctw04-03.pdf>

[MARC Formats and Standards Description]. (MARC Standards Office, Library of Congress, 2003) <http://www.loc.gov/marc/>

Memorandum of Understanding (MOU) Between the Government Printing Office and the National Archives and Records Administration, August 12, 2003,
<http://www.gpoaccess.gov/about/naramemofinal.pdf>

Metadata Object Description Schema Official Web Site. (Library of Congress, 2004)
<http://www.loc.gov/standards/mods/>

METS: An Overview & Tutorial (Library of Congress, 2003)
<http://www.loc.gov/standards/mets/METSOverview.html>

The National Bibliography of U.S. Government Information: Initial Planning Statement, June 18, 2004 www.gpoaccess.gov/about/reports/natbib0604.pdf

ONIX for Books, Product Information Message, Release 2.1, and related ONIX information,
<http://www.editeur.org/>

Strategic Vision for the 21st Century, U.S. Government Printing Office, December 1, 2004,
http://www.main.gpo.gov/pub_print/STRATEGICPLAN.html

Superseded List, 2002,
http://www.access.gpo.gov/su_docs/fdlp/pubs/suplist/index.html

Template for Standard NAL Metadata. (NAL Metadata Task Force, 2001)
<http://www.nal.usda.gov/cataloging/TEMPLATE2.pdf>

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 17, Section 1710, Index of documents: number and distribution,
<http://www.gpoaccess.gov/uscode/index.html>

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 19, Section 1711, Catalog of Government publications,
<http://www.gpoaccess.gov/uscode/index.html>

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 19, Section 1901, Definition of Government publication, available at
<http://www.gpoaccess.gov/uscode/index.html>

United States Code, 2000 edition. Title 44, Public Printing and Documents, Chapter 19, Section 1902, Availability of Government publications ..., available at
<http://www.gpoaccess.gov/uscode/index.html>

APPENDIX 4: GUIDING PRINCIPLES

[Last Update: 6/18/04]

GPO will adhere to several statements of guiding principles regarding Federal government information dissemination, including the following:

- *GPO's Report to the Congress: Study to Identify Measures Necessary For A Successful Transition To A More Electronic Federal Depository Library Program*. Principles for Federal Government Information. U.S. Government Printing Office Publication 500.11, June 1996.
http://www.access.gpo.gov/su_docs/fdlp/pubs/study/studyhtm.html

- U.S. National Commission on Libraries and Information Science (NCLIS) Principles of Public Information. <http://www.nclis.gov/info/pripubin.html>

Of specific note are the following excerpts from the NCLIS Principles of Public Information:

- The public has the right of access to public information.
- The Federal Government should guarantee the integrity and preservation of public information, regardless of its format.
- The Federal Government should ensure a wide diversity of sources of access, private as well as governmental, to public information.
- The Federal Government should not allow cost to obstruct the people's access to public information.
- The Federal Government should guarantee the public's access to public information, regardless of where they live and work, through national networks and programs like the Federal Depository Library Program.

DRAFT